ICS 07.080
CCS A 40

中华人民共和国国家标准

GB/T 40226—2021

环境微生物宏基因组检测　高通量测序法

Detection of environmental microbial metagenome—High throughput sequencing

2021-05-21 发布　　2021-12-01 实施

国家市场监督管理总局
国家标准化管理委员会　发布

目　　次

前言 …… Ⅲ
1　范围 …… 1
2　规范性引用文件 …… 1
3　术语和定义 …… 1
4　缩略语 …… 2
5　原理 …… 2
6　试验条件 …… 2
7　试剂 …… 2
8　仪器设备 …… 2
9　样品 …… 3
10　试验步骤 …… 3
11　试验报告 …… 4
12　质量控制 …… 5
附录 A（规范性）　粪便样品采集、保存、运输方法 …… 6
附录 B（规范性）　土壤样品采集、保存、运输方法 …… 7
附录 C（规范性）　水体样品采集、保存、运输方法 …… 8
附录 D（资料性）　样品信息单 …… 9
附录 E（资料性）　DNA 样品的完整性图谱和质量级别分类 …… 10
附录 F（资料性）　参考数据库 …… 11
参考文献 …… 12

前　言

本文件按照GB/T 1.1—2020《标准化工作导则　第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由全国生化检测标准化技术委员会(SAC/TC 387)提出并归口。

本文件起草单位：深圳华大生命科学研究院、深圳华大基因科技有限公司、深圳市生命科技产学研资联盟、中国测试技术研究院生物研究所、深圳华大临床检验中心。

本文件主要起草人：陈冰、肖亮、钟焕姿、李俊桦、李倩一、贾慧珏、姜华艳、周李华、唐美芳、吴昊、李陶莎、周媛、钟宏彬。

环境微生物宏基因组检测　高通量测序法

1　范围

本文件描述了采用高通量测序技术进行环境微生物宏基因组检测的术语和定义、原理、试验条件、试剂、仪器设备、样品、试验步骤、试验报告和质量控制要求。

本文件适用于运用高通量测序法进行环境微生物宏基因组的检测。

2　规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 6682　分析试验室用水规格和试验方法

GB 19489　实验室　生物安全通用要求

GB/T 35537—2017　高通量基因测序结果评价要求

3　术语和定义

下列术语和定义适用于本文件。

3.1

环境　environment

相对某项中心事物的周围世界。

注：在本文件中特指人体（动物）环境与自然环境。人体（动物）环境包括但不限于口腔、皮肤、生殖道、肠道等；自然环境包括但不限于空气、水体、土壤、沉积物等。

3.2

相对丰度　relative abundance

某一特定种类微生物在环境微生物群落中所占的相对比例。

注：通常以百分比表示。

3.3

微生物宏基因组　microbial metagenome

以特定环境中整个微生物群落作为研究对象，不分离培养，直接提取得到的所有微生物基因组DNA。

注：可用于分析其遗传信息、物种分类、系统进化、基因功能及代谢网络等。

3.4

碱基识别质量　quality of base calling

评价碱基准确识别的概率。

注：简称为Q，通常以数值表示。碱基识别质量值与碱基识别错误率负相关，二者遵循对数函数关系。碱基识别质量值越高，错误率越低。

3.5

Q20

测序数据中，碱基识别质量值为20的碱基识别准确率为99%，或错误率为1%。

3.6

Q30

测序数据中，碱基识别质量值为30的碱基识别准确率为99.9%，或错误率为0.1%。

4 缩略语

下列缩略语适用于本文件。

DNA：脱氧核糖核酸(deoxyribonucleic acid)

ID：识别码(identifier)

PCR：聚合酶链式反应(polymerase chain reaction)

RNA：核糖核酸(ribonucleic acid)

5 原理

运用高通量测序技术对环境样品中微生物基因组DNA进行测序，再将测序结果与现有数据库中序列进行比对，从而检测样品中所含微生物种类和相对丰度。

6 试验条件

6.1 实验室设施和设备要求应符合GB 19489的规定。

6.2 实验室用水应符合GB/T 6682的规定。

6.3 实验室应做到防止污染，应根据不同工作内容划分独立区域，并有明显标志，如试剂储备和准备区、样品制备区、扩增区等，各区域间应避免交叉污染。

7 试剂

7.1 DNA提取试剂。

7.2 PCR产物纯化试剂。

7.3 文库制备试剂。

7.4 高通量测序试剂。

8 仪器设备

8.1 高通量测序仪。

8.2 PCR仪。

8.3 冷冻离心机：最大离心力12 000g以上。

8.4 微量移液器。

8.5 紫外分光光度计。

8.6 荧光定量仪。

8.7 电泳仪。
8.8 凝胶成像系统。
8.9 水浴锅。
8.10 低温冰箱：−20 ℃和−80 ℃。

9 样品

9.1 样品采集、保存和运输

9.1.1 样品采集应经过伦理审查和生物安全评价审查。
9.1.2 应即刻采样，减少样品暴露于空气中的时间，同时应避免样品被其他物质污染。
9.1.3 应根据环境样品类型等情况选择相应的采样方法以保证样品中微生物群落的代表性和准确性。粪便样品采集、保存和运输方法按附录 A，土壤样品采集、保存和运输方法按附录 B，水体样品采集、保存和运输方法按附录 C 执行。

9.2 样品接收与处理

应核对样品编号，记录样品信息，检查样品状态，避免密封不严导致泄漏或污染等，否则该样品应废弃，重新采样。应记录样品信息（见附录 D），以保证样品的可追溯性。

10 试验步骤

10.1 DNA 提取

10.1.1 原理

将环境样品悬浮于裂解缓冲液中，通过抽提等步骤提取微生物基因组 DNA，充分去除样品中的蛋白质、脂类、多糖、RNA 等杂质。宜选用手工提取方法或相应 DNA 提取试剂盒。

10.1.2 DNA 样品浓度和完整性检测

环境样品提取微生物基因组 DNA 后，应使用荧光定量仪检测 DNA 样品浓度，使用琼脂糖凝胶电泳法检测 DNA 样品完整性。宜综合 DNA 样品浓度及完整性，判断 DNA 样品质量级别，判断依据见附录 E。

10.2 文库构建与高通量测序

10.2.1 应使用合适的 DNA 样品和文库构建试剂进行文库构建。进行文库构建的 DNA 样品类别选择标准按表 1。

表 1 进行文库构建的 DNA 样品类别选择标准

DNA 样品类别	是否满足文库构建
A 类	满足文库构建要求，DNA 总量≥1.0 μg
B 类	满足文库构建要求，DNA 总量≥0.5 μg，且＜1.0 μg
C 类	不完全满足文库构建要求，可以尝试进行该步骤，但不保证测序质量。 建议重新采样，重新提取样品微生物基因组 DNA

10.2.2 应使用合适的文库和高通量测序试剂进行高通量测序。高通量测序按 GB/T 35537—2017 的附录 A 进行。

10.3 数据处理

10.3.1 数据质控

10.3.1.1 质控步骤

样品经过高通量测序得到的原始数据，应进行质量控制，去除含接头或低质量的序列、外源或宿主基因组序列，再进行后续生物信息学分析。

10.3.1.2 碱基识别质量值要求

测序完成后，应进行 Q20、Q30 的统计和评估。每个 DNA 样品可用数据对应的碱基识别质量值应符合如下要求：

——大于 Q20 的碱基比例≥90%；

——大于 Q30 的碱基比例≥80%。

10.3.1.3 可用数据量要求

可用测序数据量应达到声明目标物种可从样品中检测到的最小测序数据量。每个样品测序生成的可用高质量序列数目应大于 1×10^7 条。

注：序列数目为单端测序序列条数，或双端测序序列对数。

10.3.2 物种注释

数据质控后对样品数据进行物种注释。除特殊方法外，对于已知物种，宜使用数据库中序列相似度 95%以上，覆盖度 90%以上的物种信息进行注释。数据库见附录 F。

10.3.3 序列拼接与组装

数据质控后对样品数据可进行序列拼接与组装。使用组装软件将序列进行拼接，得到更长的叠连群，将这些叠连群片段聚集起来，与参考数据库进行比对，得到新的参考基因集。数据库见附录 F。

10.3.4 微生物物种相对丰度计算

数据质控后对样品数据进行物种相对丰度计算，应对所使用的相对丰度计算方法进行记录。除特殊方法外，宜将可用高质量测序数据比对到组装好的参考基因集或合适的数据库上，按相似度 95%以上，覆盖度 90%以上进行统计，得到基因水平上的相对丰度分布情况，再将注释到同一物种分类水平的基因序列相对丰度进行叠加，得到该物种的相对丰度结果。

11 试验报告

试验报告应包含环境样品中微生物群落的物种组成及各成分相对丰度，可增加个性化功能分析部分，并应至少包括下列内容：

a) 检测机构名称、联系方式和地址、检测员、样品接收时间、样品检测时间及报告时间；

b) 测序平台与测序策略；

c) 数据分析对应使用的注释方法、软件、参数、流程和数据库版本等；

d) 样品物种组成及各相应丰度，宜选择多元展现方式。

12 质量控制

12.1 宜使用标准物质进行质量控制，评估定性与定量检测结果的准确性。

12.2 在描述结果时应注明检测结果的局限性，说明非本文件规定的可能影响检测结果的所有操作细节。

附 录 A
（规范性）
粪便样品采集、保存、运输方法

A.1 粪便杯采集法

A.1.1 采集

应使用洁净的器皿收集粪便样品，确认器皿中无水、尿液或其他分泌物（如经血）等污染。排便后即刻使用一次性粪便杯配套取样勺对粪便样品中部取样，将粪便样品从洁净器皿转移至粪便杯，立刻旋紧盖子。每个粪便杯中转移至少 2 cm^3 大小的粪便样品。

A.1.2 保存与运输

A.1.2.1 粪便杯采集粪便样品后，应立即置于干冰或－80 ℃保存。若不能即刻转移，可暂存于≤4 ℃条件下（如冰盒），30 min 内转移至干冰或－80 ℃保存。

A.1.2.2 应在干冰条件下进行运输。保存及运输过程应避免反复冻融。

A.2 含稳定液自采样套装采集法

A.2.1 采集

应使用洁净的器皿收集粪便，确认器皿中无水、尿液或其他分泌物（如经血）等污染，按照自采样套装操作要求取中段粪便，将采集的粪便样品迅速转移至含有稳定剂的保存管中，使稳定液完全浸没粪便样品，盖紧管盖。

A.2.2 保存与运输

含稳定液保存管采样后，该保存管按操作说明可于常温保存和运输，避免阳光直射。常温放置时间应不超过自采样套装操作说明限定的常温保存时间范围。对将超出自采样套装操作说明限定时间的常温粪便保存管需转移至－80 ℃冻存，应在干冰条件下进行运输。保存及运输过程应避免反复冻融。

附 录 B
（规范性）
土壤样品采集、保存、运输方法

B.1 采集

宜取土壤表层 5 cm～10 cm 处土壤，如果土壤有翻动，应更深处采样，避免空气微生物污染。采样区内设置至少 3 个重复采样点，保证所取样本对所在地域的代表性。每个点取样量应一致（≥5 g），去除土样里植物根系或砾石，将重复土样混合均匀，做好标记，装入灭菌封口聚乙烯袋或其他无菌容器。

B.2 保存与运输

B.2.1 采集样品后，应立即置于干冰或－80 ℃保存。若不能即刻转移，可暂存于≤4 ℃条件下（如冰盒），30 min 内转移至干冰或－80 ℃保存。

B.2.2 应在干冰条件下进行运输。保存及运输过程应避免反复冻融。如需常温保存或运输，宜将样品浸润于稳定剂中。

附 录 C
（规范性）
水体样品采集、保存、运输方法

C.1 采集

C.1.1 根据水体不同的浊度，应使用无菌器材采集 4 L～200 L 不等的水样。采集好的水样需要通过滤膜进行过滤，可选择不同孔径大小的滤膜。

示例：20 μm、3 μm、0.8 μm、0.4 μm、0.2 μm、0.1 μm 等孔径滤膜。

C.1.2 浊度较大水样应先静置，分离悬浮颗粒；再使用 20 μm 孔径滤膜预过滤，再选择小孔径滤膜进行过滤。浊度较小清凉水样，可选择小孔径滤膜直接过滤。

C.2 保存与运输

C.2.1 样品最后一次过滤后的滤膜应放置于无菌容器，并将无菌容器立即置于干冰或－80 ℃进行保存。若不能即刻转移，可暂存于≤4 ℃条件下（如冰盒），30 min 内应转移至干冰或－80 ℃保存。

C.2.2 应在干冰条件下进行运输。保存及运输过程中应避免反复冻融。如常温保存或运输，宜将样品浸润于稳定剂中。

附 录 D
（资料性）
样品信息单

样本信息单记录见表 D.1。

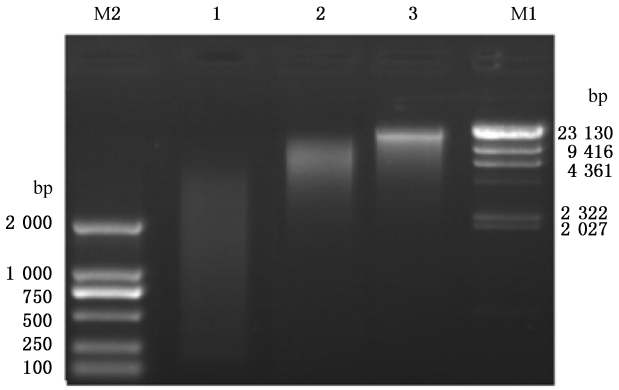
表 D.1 样本信息单

<table>
<tr><td rowspan="3">运输信息</td><td colspan="4">运输公司：_____ 快递单号：_____ 城市：_____
样品寄件人：_____ 样品寄件人联系电话：_____
运输状态：□ 干冰 □ 冰袋 □ 常温 □ 其他：_____</td></tr>
<tr><td>样品联系人 *</td><td></td><td>样品联系人电话</td><td></td></tr>
<tr><td>样品联系人邮箱 *</td><td></td><td>样品联系人单位</td><td></td></tr>
<tr><td colspan="2">有无传染性、致病性 *</td><td colspan="3">□ 无 □ 有（致病传染性说明：_____） □ 未知</td></tr>
<tr><td colspan="2">实验室要求 *</td><td colspan="3">□ 普通实验室 □ P2 实验室（生物安全柜）</td></tr>
<tr><td colspan="2">详细提取方案 *</td><td colspan="3"></td></tr>
<tr><td>样品信息</td><td>信息栏</td><td colspan="3">填写标准</td></tr>
<tr><td rowspan="3">样品名称</td><td rowspan="3"></td><td colspan="3">1. 样品名称必需唯一化以区分其他样本</td></tr>
<tr><td colspan="3">2. 样品名称可由字母及数字组成；不能包含“_”“-”“＃”“%”“/”“&”和“＊”等非常规符号</td></tr>
<tr><td colspan="3">3. 人体相关样品名称不能包含任何可识别个体身份的信息</td></tr>
<tr><td>样品状态</td><td></td><td colspan="3">样本状态包括：DNA、速冻组织（速冻粪便、速冻土壤、速冻水体等）、含稳定剂组织等</td></tr>
<tr><td>分类单元 ID</td><td></td><td colspan="3">按照 NCBI taxon ID 提供样品分类单元信息。例如：410658 代表土壤宏基因组样品；749906 代表肠道宏基因组样品；412755 代表海洋沉积物宏基因组样品；408172 代表海洋水体宏基因组样品；449393 代表淡水宏基因组样品。
NCBI 无记录 ID 的样品类型归为 408169</td></tr>
<tr><td>样品环境名称</td><td></td><td colspan="3">样本环境名称必需与上述分类单元 ID 一致。包括土壤宏基因组、肠道宏基因组、海洋沉积物宏基因组、海洋水体宏基因组等</td></tr>
<tr><td>宿主名称
（仅针对宿主相关样本，含中文名与拉丁名）</td><td></td><td colspan="3">宿主名称：人（Homo sapiens）、猪（Sus scrofa）</td></tr>
<tr><td>采集时间</td><td></td><td colspan="3">样本采集时间，按照年-月-日“yyyy-mm-dd”规范记录，对无法精确日期的样本可精确到月份或年份。例如：2017-01-23，2017-01 和 2017</td></tr>
<tr><td>采集地点</td><td></td><td colspan="3">采集地点信息需包括国家和省份信息（详细位置信息由经纬度提供）。例如：中国广东省</td></tr>
<tr><td>采集经度</td><td></td><td colspan="3">采集地点经度，正值代表东经，负值代表西经。
例如：40.743 和－10.530</td></tr>
<tr><td>采集纬度</td><td></td><td colspan="3">采集地点纬度，正值代表北纬，负值代表南纬。
例如：18.580 和－89.122</td></tr>
<tr><td colspan="5">注 1：样本信息标准参考国际基因组学标准联盟的宏基因组序列最小信息（MIMS）标准。
注 2：“＊”栏目为必填项。</td></tr>
</table>

附 录 E
（资料性）
DNA 样品的完整性图谱和质量级别分类

E.1 DNA 完整性判断

使用琼脂糖凝胶电泳法检测基因组 DNA 样品的完整性，进行完整性判断的电泳图谱见图 E.1。



标引序号说明：

1 ——重度降解：电泳图中对应样品的泳道无明显主带，拖尾严重；

2 ——中度降解：电泳图中对应样品的泳道可见主带，略有拖尾；

3 ——轻微降解：电泳图中对应样品的泳道可见主带，略有拖尾；

M1——DNA 标准品 1；

M2——DNA 标准品 2；

bp ——碱基对(base pair)。

图 E.1 基因组 DNA 样品降解程度示意图

E.2 DNA 质量级别分类

结合基因组 DNA 样品浓度和完整性，对 DNA 样品质量级别进行分类，分类标准见表 E.1。

表 E.1 DNA 样品质量级别分类表

样品类型	检测结果			判定结论
基因组 DNA	$m \geqslant 1.0\ \mu g$	$\rho \geqslant 10\ ng/\mu L$	无降解或轻微降解	A 类
	$0.5\ \mu g \leqslant m < 1.0\ \mu g$			B 类
	$m < 0.5\ \mu g$	$\rho \leqslant 10\ ng/\mu L$	中度降解或重度降解	C 类

注 1：m——指 DNA 总量。

注 2：ρ——指 DNA 质量浓度。

注 3：A 类样品同时满足 $m \geqslant 1.0\ \mu g$，$\rho \geqslant 10.0\ ng/\mu L$，无降解或轻微降解三个条件，合格。

注 4：B 类样品同时满足 $0.5\ \mu g \leqslant m < 1.0\ \mu g$，$\rho \geqslant 10.0\ ng/\mu L$，无降解或轻度降解三个条件，合格。

注 5：C 类样品满足 $m < 0.5\ \mu g$，$\rho \leqslant 10.0\ ng/\mu L$，中度降解或重度降解三个条件其中之一或以上，不完全合格。

附 录 F
（资料性）
参考数据库

数据处理过程可使用如下参考数据库：

a) 国家生物信息中心(CNCB)/国家基因组科学数据中心(NGDC)的基因组数据库
https://bigd.big.ac.cn/gwh/

b) 国家微生物科学数据中心(NMDC)的微生物宏基因组数据库
https://gcmeta.wdcm.org/

c) 国家微生物科学数据中心(NMDC)的全球微生物菌种目录数据库
http://gcm.wdcm.org/

d) 中国国家基因库生命大数据平台的微生物数据库
https://db.cngb.org/datamart/microbe

e) 美国国立生物技术信息中心(NCBI)的基因银行数据库
https://www.ncbi.nlm.nih.gov/genbank

f) 美国国立生物技术信息中心(NCBI)的参考序列数据库
https://www.ncbi.nlm.nih.gov/refseq

g) 美国国立生物技术信息中心(NCBI)的微生物基因组数据库
http://www.ncbi.nlm.nih.gov/genome/microbes

h) 美国国立生物技术信息中心(NCBI)的物种分类数据库
http://www.ncbi.nlm.nih.gov/taxonomy

i) 美国能源部联合基因组研究中心(DOE-JGI)的微生物基因组数据库
https://img.jgi.doe.gov/

j) 美国能源部联合基因组研究中心(DOE-JGI)的细菌核糖体 RNA 基因序列数据库
http://greengenes.secondgenome.com

k) 欧洲生物信息研究所(EBI)的核酸数据库
http://www.ebi.ac.uk/embl/

l) 欧洲生物信息研究所(EBI)的蛋白数据库
https://www.uniprot.org/

m) 斯坦福国际研究所(SRI international)的代谢组学数据库
https://metacyc.org/

n) 日本京都大学的基因与基因组百科全书
https://www.genome.jp/kegg/

o) 日本国立遗传学研究所(NIG)的核酸数据库
http://www.ddbj.nig.ac.jp

p) 加拿大麦克马斯特大学的抗性基因数据库
https://card.mcmaster.ca/

q) 全球海洋微生物数据库(TARA Oceans)
https://www.ocean-microbiome.org/

参 考 文 献

[1] Crits-christoph, A., Diamond, S., Butterfield, C. N., Thomas, B. C. & Banfield, J. F. Novel soil bacteria possess diverse genes for secondary metabolite biosynthesis. Nature 558, 440-444 (2018).

[2] Fierer, N. et al. Comparative metagenomic, phylogenetic and physiological analyses of soil microbial communities across nitrogen gradients. ISME J. 6, 1007-1017 (2012).

[3] Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J. & Goodman, R. M. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chem. Biol. 5, R245-R249 (1998).

[4] Lauber, C. L., Zhou, N., Gordon, J. I. & Knight, R. Effect of storage conditions on the assessment of bacterial community structure in soil and humann-associated samples. FEMS Microbiol. Rev. 307, 80-86 (2010).

[5] Methé, B. A. et al. A framework for human microbiome research. Nature 486, 215-221 (2012).

[6] Pesant, S. et al. Open science resources for the discovery and analysis of Tara Oceans data. Sci. Data 2, 1-16 (2015).

[7] Qin, J. et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature 490, 55-60 (2012).

[8] Renzo, K. et al. A Standard MIGS/MIMS compliant XML schema: toward the development of the Genomic Contextual Data Markup Language (GCDML). OMICS 12, 115-121 (2008).

[9] Song, S. J., Amir, A., Metcalf, J. L. & Amato, K. R. Preservation Methods Differ in Fecal Microbiome Stability, Affecting Suitability for Field Studies. mSystems. 1, 1-12 (2016).

[10] Thompson, L. R. et al. A communal catalogue reveals Earth's multiscale microbial diversity. Nature 551, 457-463 (2017).

[11] Wesolowska-andersen, A. et al. Choice of bacterial DNA extraction method from fecal material influences community structure as evaluated by metagenomic analysis. Microbiome 1-11 (2014).